NPS55-88-008

NAVAL POSTGRADUATE SCHOOL

Monterey, California

SYSTEM AVAILABILITY: TIME DEPENDENCE
AND STATISTICAL INFERENCE BY
(SEMI)NON-PARAMETRIC METHODS

DONALD P. GAVER
PATRICIA A. JACOBS

AUGUST 1988

Prepared for:
Office of Naval Research
Arlington, VA 22217

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral R. C. Austin
Superintendent

H. Shull
Provost

The work reported herein was supported in part with funds provided by the Office of Naval Research.

This report was prepared by:

ASSIFIED
CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

ORT SECURITY CLASSIFICATION SSIFIED	1b RESTRICTIVE MARKINGS
RITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
ASSIFICATION/DOWNGRADING SCHEDULE	
RMING ORGANIZATION REPORT NUMBER(S) -88-008	5 MONITORING ORGANIZATION REPORT NUMBER(S)
E OF PERFORMING ORGANIZATION l Postgraduate School — 6b OFFICE SYMBOL (If applicable) Code 55	7a NAME OF MONITORING ORGANIZATION
RESS (City, State, and ZIP Code) erey, CA 93943-5000	7b. ADDRESS (City, State, and ZIP Code)
E OF FUNDING/SPONSORING ANIZATION e of Naval Research — 8b. OFFICE SYMBOL (If applicable) Code 430	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N0001487WR24005
RESS (City, State, and ZIP Code) gton, VA 22217	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61153N; PROJECT NO RR014-05-01; TASK NO 4114531-02; WORK UNIT ACCESSION NO.

(Include Security Classification) SYSTEM AVAILABILITY: TIME DEPENDENCE AND STATISTICAL
ENCE BY (SEMI)NON-PARAMETRIC METHODS

ONAL AUTHOR(S)
r, Donald P.; Jacobs, Patricia A.

E OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
ical	FROM ______ TO ______	1988 August	42

LEMENTARY NOTATION

COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
D	GROUP	SUB-GROUP	availability, reliability, renewal, non-parametric estimation, bootstrapping

RACT (Continue on reverse if necessary and identify by block number)

The modelling, and assessment from data, of the availability of a system is studied. Special consideration is given to availability in finite time (not steady-state or long-run), and to non-parametric estimates.

IBUTION/AVAILABILITY OF ABSTRACT LASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
E OF RESPONSIBLE INDIVIDUAL ld P. Gaver	22b TELEPHONE (Include Area Code) (408)646-2605 — 22c OFFICE SYMBOL Code 55Gv

M 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

☆ U.S. Government Printing Office 1986—606-243

SYSTEM AVAILABILITY: TIME DEPENDENCE AND STATISTICAL INFERENCE BY (SEMI)NON-PARAMETRIC METHODS.

D.P. Gaver
P.A. Jacobs

INTRODUCTION.

Component or system availability generally refers to (a) the obability that the item of concern is operable or "up" and mission-pable at a point or during a period of time, or to (b) the fraction time, or of demands if the item is on standby, for which the system up. Applied to large systems such as entire U.S. Navy ships, system ailability is referred to as readiness. At a different level, the ailability and fault-tolerance of computer systems, both government d commercial, is of concern, and has recently been extensively delled and simulated; see Goyal, Lavenberg and Trivedi[1]. The ailability concept is also relevant when discussing the safety and oductivity of commercial nuclear power plants; in that arena it is antified by probabilistic risk assessment (PRA). Related finite ate stochastic models also occur in medical studies; c.f. Cox[2], cobs[3].

Component or system availability is influenced both by the herent failure-proneness of the item and by the time and resources it kes to restore a failed item to service. Times to failure or "up nes"and to restoration or "down" times may vary considerably, and not cessarily independently, depending upon the mode of failure, the time quired to diagnose the failure including the access to (including competition for) diagnostic equipment and human skills, the availability of spare parts (logistic delays), and other factors. It is also quite conceivable that the quality of the repair activity influences future times to failure. This effect is not recognized by the usual models; but see Gaver[4] (pp. 775-800), and Thomas, Jacobs and Gaver[5]. Incorporation of "availability growth" in availability situations is a practical issue that is not frequently modelled; see our Section 2 below, however.

The present paper addresses the assessment of simple system availability when there is concern about (a) time-dependence, so demands for system performance are not necessarily when the system is in "steady-state," as is often assumed, and when (b) information about system failures and repairs is in the form of observed data so questions of statistical influence arise. The methods and models involved lean towards the semi-parametric or non-parametric; in particular we employ the empirical Laplace transform in the time-dependent scenarios of interest; non-parametric estimation in assumed steady-state situations has been studied by Gaver and Chu[6]. Other investigations, e.g. Cox[7], Gross and Harris[8] and Ascher and Feingold[9] have been overwhelmingly concerned with estimation in presumably well-specified stochastic (queueing, Markov) models; considerations of model mis-specification are seldom broached. Here we propose analytically simple approximations to time-dependent system behavior, and assess the effects of model specification ("up" and "down" time dependence) upon rates of approach to a long-run steady state as the latter are estimated from available data (assumed to be a random sample). Uncertainty assessments (confidence limits and standard errors) are furnished. More elaborate procedures involving Bayes or empirical Bayes setups that permit "strength borrowing" (in John Tukey's phrase) are not addressed here, but are agenda items.

Our paper's plan is as follows. Section 2 describes a selected group of probability models for simple system availability; no comprehensiveness is claimed. Solutions are given in terms of Laplace transforms, all of which are rendered immediately interpretable in terms of the random-time-of-demand or observation paradigm, described originally and applied to transform inversion in Gaver[10]. Section 3 describes and "fits" the simple exponential-approach-to-steady-state model used for representing time-dependent behavior; see Odoni and Roth[13] and Gaver and Jacobs[11] where such an idea was used to represent time-dependent queueing behavior. Section 4 introduces issues of estimating time-dependent availability where only statistical data is at hand. Section 5 presents a variety of numerical illustrations.

2. TIME-DEPENDENT AVAILABILITY: MODELS, TRANSFORMS, AND THE RANDOM OBSERVATION (ROBS) INTERPRETATION.

Consider a system that is active or operative ("up") until failur occurs, after which a "down" period occurs, at the termination of whic a new up period begins, and so on. Suppose the system is to be responsive to a randomly appearing demand; what is the probability tha it will respond properly, or be available, at the time of the demand? In the military arena the device might be a surveillance (radar, sonar or communications system. In the commercial nuclear power context it might be a coolant pump.

There are many versions of this problem. The simplest and most traditional imagines that the system operates continually between failures, and failure is instantly detected upon occurrence. Up times may be modelled as iid random variables, either directly or indirectly If the system is redundant a new up time begins at the instant an old one terminates, with repair off-line. Down times may also be iid. However, there are many plausible exceptions to these scenarios, a few of which are considered next.

Here are various sample model formulations; the list is by no means complete. In all of these the system is up at $t = 0$; modifications are straightforward.

<u>Model 1</u>: up times are iid $\{U_i\}$, with df $F_U(x)$; down times are iid $\{D_i\}$, d.f. $F_D(y)$. $A(t) = P\{\text{System up at } t | \text{up time starting at } t = 0\}$. Then, by a simple backward renewal argument, $A(t)$ satisfies the integral equation

$$A(t) = \bar{F}_U(t) + \int_0^t A(t-v)F_C(dv) \tag{2.1}$$

where $F_C(t) = F_U * F_D(t)$, the convolution. Laplace-transforming,

$$\hat{A}(s) = \int_0^\infty e^{-st} A(t)dt = \frac{1 - \hat{F}_U(s)}{s[1 - \hat{F}_C(s)]} . \tag{2.2}$$

If the item is observed or demanded at random time $T \sim \exp(s)$ then

$$E[A(T)] = \int_0^\infty se^{-st} A(t)dt = \frac{1 - \hat{F}_U(s)}{1 - F_C(s)} \tag{2.3}$$

and since $E[T] = 1/s = \tau$, the probability of being up upon demand is

$$a_0(\tau) = E[A(T)] = \frac{1 - \hat{F}_U(\tau^{-1})}{1 - \hat{F}_C(\tau^{-1})} \tag{2.4}$$

$$\rightarrow \quad \frac{E[U]}{E[C]} = \frac{E[U]}{E[U] + E[D]} \quad \text{as} \quad \tau \rightarrow \infty \tag{2.5}$$

by Tauberian/Abelian results; see Feller (1966). One can view (2.4) as the availability under random demand or observation; ROBS for short. To emphasize the fact that <u>exponential</u> random observation is involved we utilize EROBS. The last formula, written here as

$$a_0(\infty) = \frac{E[U]}{E[U] + E[D]} \tag{2.6}$$

is the widely-applied <u>long-run availability</u>. It may well be inappropriate in many of the contexts in which it is applied.

Note, though, that

$$a_0(\tau) = \frac{1 - \hat{F}_U(\tau^{-1})}{1 - \hat{F}_C(\tau^{-1})} \tag{2.7}$$

has a definite interpretation for <u>all</u> τ. It is usually easily compute and interpreted whenever transforms of U and D (or C) exist. As will be seen, it can be estimated from data under many circumstances--even when the model selected has either E[U] or E[D] infinite and (2.6) becomes uninteresting.

Here are several additional models whose transforms can be directly written down, and directly interpreted under EROBS demands.

<u>Model 2</u>: The initial up time U_0 has df F_{U_0}; $\{D_i + U_i \equiv C_i, i = 1,2,...\}$ are iid and independent of U_0. This attempts to model a situation in which the quality of maintenance during a down time affects the distribution of the next up time, so these times are dependent; the influence stops at that point, however. Let $\overline{A}_D(t) =$ P{system not up (down) at t|down time starts at t = 0}. Then

$$\overline{A}_D(t) = \overline{F}_D(t) + \int_0^t \overline{A}_D(t-r)F_C(dr)$$

and

$$A(t) = \begin{cases} 1 & \text{if } U_0 > t, \\ A_D(t-U_0) = 1 - \overline{A}_D(t-U_0) & \text{if } U_0 \le t \end{cases}$$

so

$$A(t) = \overline{F}_{U_0}(t) + \int_0^t [1 - \overline{A}_D(t-r)]F_{U_0}(dr) = 1 - \int_0^t \overline{A}_D(t-r)F_{U_0}(dr).$$

Transforming gives

$$\hat{A}(s) = \frac{1}{s}\left[\frac{1 - \hat{F}_C(s) - \hat{F}_{U_0}(s) + \hat{F}_{U_0}(s)\hat{F}_D(s)}{1 - \hat{F}_C(s)}\right] \tag{2.8}$$

from which the observational probability $a_0(\tau)$ can be immediately written down: $a_0(\tau) = \tau^{-1}A(\tau^{-1})$. Models that propose that the up time can positively influence the duration of the following down time have been formulated; see Gaver[4].

<u>Model 3</u>: (Random changepoint model; "reliability growth"). Suppose the system fails according to up times $\{U_{i1},\ i = 1,2,\ldots I\}$, iid $F_1(u)$ and thereafter $\{U_{i2},\ i = I+1, I+2, \ldots\}$, iid $F_2(u)$; $I \sim p_i$ is the changepoint and is associated with diagnostic-repair activity. Let $F_D(u)$ describe the iid down times $\{D_i,\ i = 1,2,\ldots\}$. Suppose the system starts operation as a Type 1 up time, transitioning after a random number of down times to Type 2. Let $A_{1j}(t)$ denote the probability that the system is up and in state j at time t given it is up at time 0, i.e. t is contained in an up time of Type j. Then, using backward renewal arguments,

$$A_{11}(t) = \bar{F}_{1U}(t) + (1-p_1)\int_0^t A_{11}(t-r)F_{1C}(dr) ,$$

$$A_{22}(t) = \bar{F}_{2U}(t) + \int_0^t A_{22}(t-r)F_{2C}(dr) , \tag{2.9}$$

and

$$A_{12}(t) = \int_0^t A_{22}(t-r)G(dr) ,$$

where

$$G(r) = \sum_{i=1}^{\infty} p_i F_{1C}^{i*}(r) ;$$

as before $F_{iC}(r)$ refers to the df of a cycle time $U_{ji} + D_j$. Transforming immediately produces a solution:

$$\hat{A}_{11}(s) = \frac{1}{s} \frac{1 - \hat{F}_{1U}(s)}{1 - (1-p_1)\hat{F}_{1C}(s)}$$

$$\hat{A}_{22}(s) = \frac{1}{s} \frac{1 - \hat{F}_{2U}(s)}{1 - \hat{F}_{2C}(s)} \tag{2.10}$$

$$\hat{A}_{12}(s) = \hat{A}_{22}\hat{G}(s) ,$$

and

$$\hat{G}(s) = \sum_{i=0}^{\infty} p_i\left(\hat{F}_{1C}(s)\right)^i = p\left(\hat{F}_{1C}(s)\right) ,$$

this last being the generating function of the number of repairs to transition rv, I. Assemble to obtain the transform of the probability of being up at t:

$$\hat{A}_1(s) = \int_0^t e^{-st}[A_{11}(t) + A_{12}(t)]dt = \hat{A}_{11}(s) + \hat{A}_{12}(s)$$

$$= \frac{1}{s}\left[\frac{1 - \hat{F}_{1U}(s)}{1 - (1-p_1)\hat{F}_{1C}(s)} + p\left(\hat{F}_{1C}(s)\right) \cdot \frac{1 - \hat{F}_{2U}(s)}{1 - \hat{F}_{2C}(s)}\right] . \tag{2.11}$$

Consequently the observational probability

$$a_0(\tau) = E[A(T)] = \left[\frac{1 - \hat{F}_{1U}(1/\tau)}{1 - (1-p_1)\hat{F}_{1C}(1/\tau)} + p\left(\hat{F}_{1C}(1/\tau)\right) \frac{1 - \hat{F}_{2U}(1/\tau)}{1 - \hat{F}_{2C}(1/\tau)}\right] \tag{2.12}$$

$$\rightarrow \frac{E[U_2]}{E[U_2] + E[D]} \quad \text{as} \quad \tau \rightarrow \infty$$

but the latter limiting formula provides no information about the availability at early times, possibly before change has taken place. The EROBS random demand formulation at least allows some information to be obtained in a very direct and meaningful manner by just evaluating the transform itself numerically.

Model 4: (Markovian changepoints). It is possible that a system alternates slowly between failure states, either temporarily or forever. The random motion may be in response to occasional changes in maintenance practices, to debugging (reliability growth), or to ageing (reliability decay) and subsequent replacement. To illustrate, let there be just two failure distribution states, as in Model 3, but let $\{p_{ij};\ i,j = 1,2\}$ be the transition probabilities of a Markov chain that governs jumps between them. Suppose the system starts in state 1, at the beginning of an up time $U_i \sim F_{iU}(r)$, $i = 1,2$. Then, letting $A_{ij}(t)$ be the probability the system is up and in state j at time t given it just started an up time at time 0 and was in state i,

$$A_{11}(t) = \overline{F}_{1U}(t) + p_{11}\int_0^t A_{11}(t-r)F_{1C}(dr) + p_{12}\int_0^t A_{21}(t-r)F_{1C}(dr),$$

$$A_{12}(t) = p_{12}\int_0^t A_{22}(t-r)F_{1C}(dr) + p_{11}\int_0^t A_{12}(t-r)F_{1C}(dr), \tag{2.13}$$

$$A_{22}(t) = \overline{F}_{2U}(t) + p_{22}\int_0^t A_{22}(t-r)F_{2C}(dr) + p_{21}\int_0^t A_{12}(t-r)F_{2C}(dr),$$

$$A_{21}(t) = p_{21}\int_0^t A_{11}(t-r)F_{2C}(dr) + p_{22}\int_0^t A_{21}(t-r)F_{2C}(dr).$$

Transforming yields these equations,

$$\hat{A}_{11}(s) = \frac{1 - \hat{F}_{1U}(s)}{s} + p_{11}\hat{A}_{11}(s)\hat{F}_{1C}(s) + p_{12}\hat{A}_{21}(s)\hat{F}_{1C}(s) \tag{2.14,a}$$

$$\hat{A}_{12}(s) = p_{12}\hat{A}_{22}(s)\hat{F}_{1C}(s) + p_{11}\hat{A}_{12}(s)\hat{F}_{1C}(s) \tag{2.14,b}$$

$$\hat{A}_{22}(s) = \frac{1 - \hat{F}_{2U}(s)}{s} + p_{22}\hat{A}_{22}(s)\hat{F}_{2C}(s) + p_{21}\hat{A}_{12}(s)\hat{F}_{2C}(s) \tag{2.14,c}$$

$$\hat{A}_{21}(s) = p_{21}\hat{A}_{11}(s)\hat{F}_{2C}(s) + p_{22}\hat{A}_{21}(s)\hat{F}_{2C}(s). \tag{2.14,d}$$

Solve (2.14,a) and (2.14,d) simultaneously to get

$$\hat{A}_{11}(s) = \frac{1}{s}\left[\frac{1 - \hat{F}_{1U}(s)}{1 - p_{11}\hat{F}_{1C}(s) - p_{12}p_{21}\hat{F}_{1C}(s)\hat{F}_{2C}(s)/\left(1-p_{22}\hat{F}_{2C}(s)\right)}\right] \tag{2.15}$$

From (2.14b,c),

$$\hat{A}_{12}(s) = \frac{1}{s}\left[\frac{1 - \hat{F}_{2U}(s)}{1-p_{22}\hat{F}_{2C}(s)-p_{21}p_{12}\hat{F}_{1C}(s)\hat{F}_{2C}(s)/\left(1-p_{11}\hat{F}_{1C}(s)\right)}\right] \cdot \left(\frac{p_{12}\hat{F}_{1C}(s)}{1-p_{11}\hat{F}_{1C}(s)}\right). \tag{2.16}$$

It follows that if $\tau = 1/s$,

$$\begin{aligned} a_1(\tau) &= s\hat{A}_{1t} = s\left[\hat{A}_{11}(s) + \hat{A}_{12}(s)\right] \\ &= \frac{1 - \hat{F}_{1U}(s)}{[1-p_{11}\hat{F}_{1C}(s)-p_{12}p_{21}\hat{F}_{1C}(s)\hat{F}_{2C}(s)/\left(1-p_{22}\hat{F}_{2C}(s)\right)} \\ &\quad + \frac{1 - \hat{F}_{2U}(s)}{1-p_{22}\hat{F}_{2C}(s)-p_{12}p_{21}\hat{F}_{1C}(s)\hat{F}_{2C}(s)/\left(1-p_{11}\hat{F}_{1C}(s)\right)} \\ &\quad \cdot \left(\frac{p_{12}\hat{F}_{1C}(s)}{1-p_{11}\hat{F}_{1C}(s)}\right). \end{aligned} \tag{2.17}$$

The above formula is manifestly opaque as it stands. Tauberian/Abelian results present the long-run behavior:

$$a_{11}(\infty) = \lim_{s \to 0} s\hat{A}_{11}(s) = \frac{E[U_1]}{E[C_1] + (p_{12}/p_{21})E[C_2]} \tag{2.18}$$

and

$$a_{12}(\infty) = \frac{E[U_2]}{E[C_2] + (p_{21}/p_{12})E[C_1]} \tag{2.19}$$

so the sum

$$a_1(\infty) = a_{11}(\infty) + a_{12}(\infty) = \frac{E[U_1]}{E[C_1] + (p_{12}/p_{21})E[C_2]} + \frac{E[U_2]}{E[C_2] + (p_{21}/p_{12})E[C_1]} \tag{2.20}$$

represents the long-run probability that the system is up. Of course $a_1(\tau)$ for τ finite is available from (2.17) and potentially provides considerably more information.

Many other such models can be constructed and "solved" by transforms. In what follows we illustrate the way in which such transform solutions can be exploited to yield time-dependent and inferential information. The intention is to provide simple and flexible approximate information rather than to utilize more exact, but also more computationally demanding, procedures.

3. TIME-DEPENDENT AVAILABILITY AT FIXED OBSERVATION TIMES (FOBS): SIMPLE EXPONENTIAL REPRESENTATIONS.

It is familiar from the theory of finite Markov chains and also from renewal theory that if a long-run or steady-state condition is reached by a stochastic model, then the nature of the approach is often essentially exponential. In particular, if, in Model 1, $U \sim \exp(\lambda)$, $D \sim \exp(\mu)$ then for $A(0) = 1$,

$$A(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}. \tag{3.1}$$

Such a time dependence is exhibited approximately by the M/G/1 queue; see Odoni and Roth[13]; Morse[14,15], and later Kielson[16], and others have assessed the relaxation times of various queueing systems. See also Gaver and Jacobs[11] for a brief discussion of the time-dependent M/G/1 queue utilizing such a representation.

It is well-understood that natural generalizations of the Markov formulation producing (3.1) can, for certain models, lead to damped oscillatory approach to a steady-state value $a(\infty)$, or to ultimate approach that is exponential modified by a negative power of t. However, in what follows we shall assume that the "error" at time t is approximately exponential:

$$e(t) = A(t) - A(\infty) \simeq \alpha e^{-\beta t}, \tag{3.2,a}$$

or, more elaborately,

$$\simeq (\alpha_1 + i\alpha_2) e^{-(r+i\theta)t} + (\alpha_1 - i\alpha_2) e^{-(r-i\theta)t} \tag{3.2,b}$$

and investigate ways of assessing parameter values from (a) perfectly specified models, wherein the distributions of U, D are presumed known, and (b) from data. This section deals with problem (a).

The motivation for considering simple exponential model (3.2) is the desire for an easily comprehended and computed assessment of time-dependent availability at a _fixed_ demand or observation time (abbreviated FOBS). In many cases encountered such an assessment (or assessments) of α and β or of α_1, α_2, r and θ can provide a useful sense of the behavior of A(t) as time progresses without the necessity of an extremely time-consuming and computer-intensive transform inversion procedure or of a symbolic inversion in terms of polynomial roots; the latter is "explicit," but hardly comprehensible in general. Simulations and numerical solutions of governing integral equations are also useful approaches, but they are generally more computationally intensive than our proposals here.

We suggest several ways of matching an exponential to A(t) by utilizing the transform $\hat{A}(s)$, presumed given.

3.1 Method 1: Least Squares, Unweighted and Weighted ("Tuned")

Begin by considering (3.2,a). If

$$e(t) = A(t) - A(\infty); \tag{3.3}$$

we wish to represent this "error" by $\tilde{e}(t) = \alpha e^{-\beta t}$. Consider

$$\Delta(\alpha,\beta) = \int_0^\infty \left(e(t)-\tilde{e}(t)\right)^2 dt = \int_0^\infty \left(e(t)-\alpha e^{-\beta t}\right)^2 dt \ , \tag{3.4}$$

the integrated squared error; the object is to minimize Δ by choice of α and β. Differentiation on α easily gives the optimizing condition

$$\alpha_0 = 2\beta \int_0^\infty e(t)e^{-\beta t}dt = 2\beta\hat{e}(\beta) = 2\beta\left[\hat{A}(\beta) - \frac{A(\infty)}{\beta}\right] ; \tag{3.5}$$

i.e. α_0 is evaluated in terms of the known Laplace transform of A(t) and of e(t), evaluated at β. To find a minimizing β it is possible to proceed by squaring the integrand of (3.4), discarding the $e^2(t)$ part, and substituting α_0 for α. Simplification leads to

$$-\Delta\Big(\alpha_0(\beta),\beta\Big) = \overline{\Delta}(\beta) = -2\beta\big[\hat{e}(\beta)\big]^2 \tag{3.6}$$

which can be readily searched for a global maximum, β_0. It is easily verified analytically that if $U \sim \exp(\lambda)$, $D \sim \exp(\mu)$ in Model 1 then $\beta_0 = \lambda + \mu$, as it should. In general the search of (3.6) must be conducted numerically.

The straightforward least-squares procedure can be <u>tuned</u> towards different time ranges by appropriate weighting. We may seek to minimize

$$\Delta(\alpha,\beta;w) = \int_0^\infty \Big(e(t)-\alpha e^{-\beta t}\Big)^2 w(t;\tau)dt \tag{3.7}$$

where

$$w(t;\tau) \approx \begin{cases} 1 & \text{near } \tau \\ 0 & \text{far from } \tau. \end{cases} \tag{3.8}$$

hen the minimizing α and β values, now $\alpha_0(\tau)$, $\beta_0(\tau)$, should lead to an xponential approximation that performs especially well near τ-- lthough perhaps less well further away. Again it is convenient to elect an analytical form for $w(t;\tau)$ that is compatible with the aplace transform, i.e. involves linear combinations of exponentials. n example is the density function of exponential order statistics; his device was used in a numerical transform inversion scheme Gaver[10]; Gaver and Jacobs[11] mention the idea in the present context).

Here is a specific example of weighting. Let

$$w(t;\tau) = 3e^{-\mu t}\cdot e^{-\mu t}\mu(1-e^{-\mu t}) = 3(e^{-2\mu t}\mu - e^{-3\mu t}\mu) \tag{3.9}$$

here $\mu = \frac{5}{6\tau}$. Then minimization leads first to

$$\frac{\partial\Delta}{\partial\alpha} = \int_0^\infty \left(e(t)-\alpha e^{-\beta t}\right)e^{-\beta t}\cdot 3\left(e^{-2\mu t}\mu - e^{-3\mu t}\mu\right)dt = 0$$

hich gives

$$\hat{e}(\beta+2\mu) - \hat{e}(\beta+3\mu) - \alpha\left(\frac{1}{2\beta+2\mu} - \frac{1}{2\beta+3\mu}\right) = 0 \tag{3.10}$$

r

$$\alpha_0(\beta) = \frac{\hat{e}(\beta+2\mu) - \hat{e}(\beta+3\mu)}{1/(2\beta+2\mu) - 1/(2\beta+3\mu)}$$

$$= [\hat{e}(\beta+2\mu) - \hat{e}(\beta+3\mu)]\ \frac{(2\beta+2\mu)(2\beta+3\mu)}{\mu}\ . \tag{3.11}$$

Squaring and omitting the term in $e^2(t)$ delivers

$$-\Delta\Big(\alpha_0(\beta),\beta\Big) = \bar{\Delta}(\beta) = (2\beta+2\mu)(2\beta+3\mu)[\hat{e}(\beta+2\mu)-\hat{e}(\beta+3\mu)]^2 , \tag{3.12}$$

which can be easily searched for a global maximum. The estimation procedure is tuned towards $\tau = \frac{5}{6\mu}$ in this case. A sharper tuning can be accomplished by a weight function that concentrates more tightly around τ than does (3.9); the latter is recognizably the density of the median of a sample of size 3 from an exponential df. If w is the density of the mth order statistic of a sample of n exponentials then its mean is $\frac{1}{\mu}\left[\frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{n-m+1}\right] \sim \frac{1}{\mu}\ln(1 - \frac{m}{n})$ and its variance is $\sim \frac{1}{\mu^2}\ \frac{m-1}{n(n-m+1)}$ if m is proportional to n. As n increases the density function of $w = w_n$ approaches a Gaussian/Normal density with variance decreasing like 1/n, which is essentially a delta function at τ as $n \to \infty$.

3.2 <u>Method 2: Derivative Matching: Exploiting Random Observation Times (ROBS)</u>.

Since it is proposed to approximate $e(t) = A(t)-A(\infty)$ by $\alpha e^{-\beta t}$, and since the transforms of both sides are known (or can be estimated from data) one may contemplate determining α and β as the solution of

$$e(\tau) = \int_0^\infty e^{-u/\tau}\big[A(u)-A(\infty)\big]\,\frac{du}{\tau} = \frac{1}{\tau}\hat{A}(1/\tau)-A(\infty)$$

$$= \int_0^\infty e^{-u/\tau}\alpha e^{-\beta u}\,\frac{du}{\tau} = \alpha\,\frac{1}{1+\beta\tau}\ ; \tag{3.13}$$

that is, the system is observed at a random exponential time having mean τ. One choice for τ is t. Obviously any α, β so found must mplicitly depend upon τ; in many situations for which the exponential s nearly correct the dependency of β and α on τ will be gentle. Here re two derivative-like prescriptions for finding $\alpha(\tau)$, $\beta(\tau)$:

a) Solve

$$\frac{1}{\alpha} + \frac{\beta}{\alpha}(\tau + \Delta) = r(\tau + \Delta) = 1/\hat{e}(\tau + \Delta)$$
$$\frac{1}{\alpha} + \frac{\beta}{\alpha}(\tau - \Delta) = r(\tau - \Delta) = 1/\hat{e}(\tau - \Delta) \tag{3.14}$$

or $\alpha = \alpha(\tau;\Delta)$, $\beta = \beta(\tau;\Delta)$.

b) Invert and analytically differentiate at an interesting τ-value

$$r(\tau) = \frac{\tau}{\hat{\bar{A}}(1/\tau) - \tau A(\infty)} = \frac{1}{\alpha} + \frac{\beta}{\alpha}\tau \tag{3.15}$$

$$r'(\tau) = \frac{\beta}{\alpha} . \tag{3.16}$$

Now solve simultaneously for $\beta(\tau)$, $\alpha(\tau)$. A good diagnostic step is o plot $\alpha(\tau)$ and $\beta(\tau)$ vs. τ; if the plots are nearly constant with τ hen the exponential provides a useful form; if the change is gradual t can potentially be represented by an empirical function, e.g. a low-rder polynomial.

It is also possible to match the exponential approximation by τ-uning, as was done by the weighting procedure of (3.7). Simply bserve the system with probability density

$$w(t;\tau) = c(n,m)(1-e^{-\mu t})^{m-1}e^{-\mu t}\mu(e^{-\mu t})^{n-m} \; ; \tag{3.17}$$

t is convenient to take n odd and m = [n/2].

3.3 Method 3: Matching Transform Means.

The following procedure represents another possibility for assessing β. Put $\alpha = 1 - A(\infty)$ and write

$$A(t) = A(\infty) + (1 - A(\infty))P(t) \tag{3.18}$$

and, denoting the transform of P by p,

$$\int_0^\infty se^{-st} A(t)dt \equiv s\hat{A}(s) = A(\infty) + (1 - A(\infty))sp(s) \quad . \tag{3.19}$$

Note that if

$$\frac{A(t) - A(\infty)}{1 - A(\infty)} \simeq e^{-\beta t}, \tag{3.20}$$

then transformation yields

$$sp(s) \simeq \frac{1}{1 + \beta(\frac{1}{s})}$$

or

$$[1 + \beta(\tfrac{1}{s})]sp(s) = 1, \tag{3.21}$$

so

$$\beta = \frac{1 - sp(s)}{p(s)} \quad . \tag{3.22}$$

To explicate the behavior of β for large t, expand (3.22) in powers of s for s small:

$$\beta = \frac{s[1 - s\hat{A}(s)]}{s\hat{A}(s) - A(\infty)} \tag{3.23}$$

$$= \frac{sE[C]\{\hat{F}_C(s) - \hat{F}_U(s)\}}{E[C][1 - \hat{F}_U(s)] - E[U][1 - \hat{F}_C(s)]}$$

$$= \frac{sE[C]\{-s(E[C]-E[U]) + \frac{s^2}{2}(E[C^2]-E[U^2])+o(s^2)\}}{E[C]\{sE[U] - \frac{s^2}{2}E[U^2]\} - E[U]\{sE[C] - \frac{s^2}{2}E[C^2]\}+o(s^2)} .$$

Thus,

$$\lim_{s\to 0} \beta(s) = \frac{2E[C]\{E[C] - E[U]\}}{E[C^2]E[U] - E[C]E[U^2]} . \tag{3.24}$$

It is anticipated that this expression is useful when t becomes large, provided a simple exponential is an appropriate approximation.

We will evaluate (3.22) for s = 1/t, t being the time of interest. It is also possible to tune this procedure to a time t as before. In particular, setting

$$\int_0^\infty w(t;\tau) \left\{\frac{A(t) - A(\infty)}{1 - A(\infty)}\right\} dt = \int_0^\infty w(t;\tau)e^{-\beta t} \tag{3.25}$$

and solving for β where, for example, w is as in (3.9). This tactic will be employed for inferential purposes in Section 5 with $w(t;\tau)$ of the form in (3.9).

4. TIME-DEPENDENT APPROACH CHARACTERISTICS ASSESSED FROM DATA.

In this section we examine situations reasonably represented as indicated previously, but for which limited information on the up and down times is available: only samples of finite sizes n_u for the up times and n_d for the down times are at hand. Furthermore it is desired to avoid using particular conventional analytical forms for the underlying distributions; instead "non-parametric" or "distribution-free" methods will be invoked. Finally, the availability at a finite time after an initial moment is of interest.

4.1 Predicting Availability Assuming EROBS: Point and Interval Estimates.

Suppose we wish to estimate the availability of a Model 1-type system at a random (exponential) time with mean τ, i.e. assuming EROBS. Data are available on the up times and down times:

$u_1, u_2, \ldots, u_{n_u}$,

$d_1, d_2, \ldots, d_{n_d}$

are the respective observations; although the order of observation is $u_1, d_1, u_2, d_2, \ldots$ etc., it will first be assumed that the u_i's and d_j's are independently sampled from fixed distributions F_U and F_D, but the latter process is otherwise unspecified.

A natural non-parametric procedure is to define the sample or empirical transforms

$$\phi_U(s) = \frac{1}{n_u} \sum_{i=1}^{n_u} e^{-su_i}, \quad \phi_D(s) = \frac{1}{n_d} \sum_{i=1}^{n_d} e^{-sd_i} \tag{4.1}$$

and to utilize these in place of the population or true transforms that appear in (2.4) as a solution to the problem posed. Thus a non-parametric estimate of the desired availability is

$$\hat{E}[A(T)] \equiv \hat{a}_0(\tau) = \frac{1 - \phi_U(\tau^{-1})}{1 - \phi_U(\tau^{-1})\phi_D(\tau^{-1})} . \tag{4.2}$$

Notice that the up and down times are conveniently assumed independent here. Procedures to deal with more complex models can and should be devised; accommodation to possible up and down data dependencies is a natural step, first by graphics and subsequently by model fitting and testing.

There are various options for estimating the sampling variability of the estimate (4.2) given the basic model. A classical procedure, adopted here, is to first note that under our assumptions both $\phi_U(s)$ and $\phi_D(s)$ are asymptotically Gaussian/Normal by the central limit theorem, for they are seen to be (modelled as) averages of independent and even <u>bounded</u> random variables e^{-sU}, e^{-sD}; $s = 1/\tau > 0$. We find easily that the corresponding random variables $\phi_U(s)$, $\phi_D(s)$ are approximately distributed as follows:

$$\begin{aligned} &\phi_U(s) \sim N\left(\hat{F}_U(s),\ \frac{\hat{F}_U(2s) - (\hat{F}_U(s))^2}{n_u}\right) ; \\ &\phi_D(s) \sim N\left(\hat{F}_D(s),\ \frac{\hat{F}_D(2s) - (\hat{F}_D(s))^2}{n_d}\right) . \end{aligned} \tag{4.3}$$

Since $\hat{a}_0(\tau)$ is a probability, confined between zero and one, it is reasonable to carry out further asymptotics on a transformation; we choose to study the logistic transformation

$$\ell(\tau) = \ln\left[\frac{\hat{a}_0(\tau)}{1-\hat{a}_0(\tau)}\right] = \ln\left[\frac{1-\phi_U(s)}{\phi_U(s)\big(1-\phi_D(s)\big)}\right] \tag{4.4}$$

$$= \ln[1-\phi_U(s)] - \ln \phi_U(s) - \ln[1-\phi_D(s)] \ .$$

Expand in Taylor series (the delta method) to obtain

$$E[\ell(\tau)] \simeq \ln\left[\frac{a_0(\tau)}{1-a_0(\tau)}\right] + \frac{1}{2}\,\frac{1-2\hat{F}_U}{\big((1-\hat{F}_U)\hat{F}_U\big)^2}\mathrm{Var}[\phi_U] + \frac{1}{2}\,\frac{\mathrm{Var}[\phi_D]}{(1-\hat{F}_D)^2}$$

$$\approx \ln\left[\frac{a_0(\tau)}{1-a_0(\tau)}\right] + \frac{1}{2}\left\{\left(\frac{1-2\phi_U}{(1-\phi_U)\phi_U)^2}\right)s_u^2/n_u + \frac{s_d^2/n_d}{(1-\phi_D)^2}\right\} , \tag{4.5}$$

where we have replaced $\hat{F}_U$ and $\hat{F}_D$ by their estimates ϕ_U and ϕ_D in the correction term. Furthermore,

$$\mathrm{Var}\big[\ell(\tau)\big] \simeq \frac{1}{\big(\hat{F}_U(1-\hat{F}_U)\big)^2}\ \mathrm{Var}[\phi_U] + \frac{1}{(1-\hat{F}_D)^2}\ \mathrm{Var}[\phi_D]$$

$$\approx \frac{s_u^2/n_u}{(\phi_U(1-\phi_U))^2} + \frac{s_d^2/n_d}{(1-\phi_D)^2} = s_\ell^2 \ . \tag{4.6}$$

We have put

$$s_u^2 = \frac{1}{n_u - 1}\sum_{i=1}^{n_u}\left(e^{-su_i} - \phi_U\right)^2$$

and (4.7)

$$s_d^2 = \frac{1}{n_d - 1}\sum_{i=1}^{n_d}\left(e^{-sd_i} - \phi_D\right)^2 \ .$$

Finally, approximate $\alpha\cdot 100\%$ confidence limits are given by

$$\ell^* + z_{\alpha/2} \cdot s_\ell \le \ln[a_0(\tau)/(1-a_0(\tau))] \le \ell^* + z_{1-\alpha/a} s_\ell \tag{4.8}$$

and

$$\ell^* = \ln\left(\frac{1-\phi_U(1/\tau)}{\phi_U(1/\tau)(1-\phi_D(1/\tau))}\right) - \frac{1}{2}\left\{\left(\frac{1-2\phi_U}{((1-\phi_U)\phi_U)^2}\right)\frac{s_u^2}{n_u} + \frac{1}{(1-\phi_D)^2}\frac{s_d^2}{n_d}\right\}, \tag{4.9}$$

while $s_\ell = \sqrt{s_\ell^2}$ from (4.6). The expression (4.8) is then inverted to provide two-sided confidence limits for the actual availability. A numerical example is provided in a later section.

4.2 Predicting Availability Semi-Parametrically at Fixed Time (According to FOBS).

Utilize the same data as that in section 4.1 but suppose we wish to estimate

$$A(t) \simeq A(\infty) + \alpha e^{-\beta t} . \tag{4.10}$$

After some experimentation it has been found that Method 3 above can be most easily and effectively adopted to estimate β and α: in (3.18) put

$$\hat{A}(\infty) = \frac{\bar{u}}{\bar{u} + \bar{d}} \tag{4.11}$$

and estimate β by evaluating the empirical version of (3.22). In the simplest case (Model 1), the empirical version of p(s) is given by

$$s\tilde{p}(s) = \frac{\dfrac{1-\phi_U(s)}{1-\phi_C(s)} - \dfrac{\bar{u}}{\bar{u}+\bar{d}}}{\dfrac{\bar{d}}{\bar{u}+\bar{d}}} \tag{4.12}$$

so the estimator of (3.22) is

$$\tilde{\beta}_e(s) = \frac{1 - s\tilde{p}(s)}{\tilde{p}(s)} . \tag{4.13}$$

Setting

$$w(t;\tau) = 3\mu(e^{-2\mu t} - e^{-3\mu t})$$

in (3.25) results in the equation

$$3\mu[\tilde{p}(2\mu) - \tilde{p}(3\mu)] = 3\mu \left[\frac{\mu}{(2\mu + \beta)(3\mu + \beta)}\right]$$

or

$$\beta^2 + 5\mu\beta + 6\mu^2 - \frac{\mu}{\tilde{p}(2\mu) - \tilde{p}(3\mu)} = 0 \tag{4.14}$$

a quadratic in β. The positive solution to (4.14), β_q, gives another estimate of β:

$$\tilde{\beta}_q = \tfrac{1}{2}[-5\mu + \sqrt{\mu^2 + 4\mu[\tilde{p}(2\mu) - \tilde{p}(3\mu)]^{-1}}] \tag{4.15}$$

where $\mu = \frac{5}{6t}$. In practice $\tilde{\beta}_q$ may occasionally be negative or imaginary, in which case a reasonable alternative is required. In our simulation, tests of the procedure defaulted to an estimate appropriate for U,D independent and exponentially distributed, i.e. $\tilde{\beta} = \left(\frac{1}{\bar{u}} + \frac{1}{\bar{d}}\right)$. However in general a non-positive $\hat{\beta}$ may indicate inappropriateness of the exponential-approach model (4.10) so a problem-dependent alternative must be sought.

The asymptotic behavior of the proposed estimates is not explored here. In the following section we describe bootstrapping (sample re-use) procedures and results for accessing sampling variability of estimates of $A(t)$ and β.

5. NUMERICAL ILLUSTRATIONS BY SIMULATION.

The procedures described earlier will now be illustrated, and to some degree tested, using simulated data.

5.1 Predicting Availability under EROBS: Illustrations

Refer to Section (4.1) for the basic approach to be utilized. Suppose that $n_u = n_d$ observations are available on iid exponentially distributed up and down time random variables; $E[U] = 1$, $E[D] = 4$. These data are to be processed utilizing expressions (4.8) and (4.9). To do so, a total of $n_s = 500$ independent simulations were carried out for the cases $n_u = n_d = n = 10$, and $n_u = n_d = n = 25$. In Table 5.1 are reported the mean values of the estimates of availability at an exponential time with various means τ. Results on 95% confidence levels are also given (average values in parentheses).

In order that the point and interval estimates behave as well as they do for small samples ($n \approx 10$) the bias correction recognizable in (4.9) is required, and the normal percent-points are best replaced by Student t percent-points with n-1 degrees of freedom. An alternative to the above procedures is to apply the jackknife; see Gaver and Chu[6]. However, the present method is perhaps more easily carried out on small computers utilizing nominal confidence levels $(1-\alpha)\cdot 100\%$. The fraction of the confidence limits covering/surrounding true availability at τ was tabulated, as were the mean upper and lower confidence limits on availability. The following table provides a summary of the results; as can be seen the coverage is close to the nominal 95%.

Table 5.1

Coverage and Mean Confidence Limits, EROBS
Nominal Confidence Level 95%
τ-Values

		0.125	1.00	10.00
	True:	(lower) 0.92 (upper)	(lower) 0.83 (upper)	(lower) 0.80 (upper)
Availability: (n = 10)	Estimated:	(0.71) 0.90 (0.96)	(0.66) 0.84 (0.93)	(0.63) 0.81 (0.91)
Coverage (%)		94	96	94
Availability: (n = 25)	Estimated	(0.82) 0.93 (0.97)	(0.74) 0.83 (0.90)	(0.70) 0.80 (0.88)
Coverage (%)		93	97	96

The figures in parentheses represent mean upper and lower confidence limits; notice that, as anticipated, these tend to move towards the true values as the sample size increases.

5.2 Predicting Availability under FOBS: Illustrations

Again we use simulation to assess the accuracy of the proposed exponential approximations given in Section 3.

It has been found experimentally that the empirical-transform-adapted least-squares approaches of Section 3.1 tend to be numerically unstable for small sample sizes; they behave well when component models are assumed perfectly specified ("known") or for very large sample sizes. On the other hand, Method 3 of Section 3.3, adapted to the empirical transform as outlined in Section 4.2, performs satisfactorily for the various cases considered and is not computationally intensive.

Simulation results for three different estimators for A(t) are reported. The estimators are all of the form

$$\hat{A}(t) = \hat{\pi} + (1 - \hat{\pi})e^{-\hat{\beta}t} \tag{5.1}$$

where

$$\hat{\pi} = \frac{\bar{u}}{\bar{u} + \bar{d}} .$$

The estimators differ in the manner in which β is estimated. The firs estimate, $\hat{A}_p(t)$, estimates β by

$$\beta_p = (\bar{u}^{-1} + \bar{d}^{-1}) . \tag{5.2}$$

Note that $\hat{A}_p(t)$ is the estimator that would be obtained if it were simply assumed that $\{U_i\}$ and $\{D_i\}$ are independent sequences of independent exponential random variables and maximum likelihood is use to obtain estimates of $A(t)$, i.e. if the simplest Markov model were automatically invoked. The second estimator, $\hat{A}_e(t)$, estimates β as in (4.13) with $s = \frac{1}{t}$. The third estimator, $\hat{A}_q(t)$, estimates β as in (4.15) with $\mu = \frac{5}{6t}$. If the estimated β's are negative then β is set equal to β_p, the MLE estimator of β if it were known that $\{U_i\}$ and $\{D_i\}$ are independent sequences of independent identically distributed exponential random variables. In our experience such pathological cases are rare.

All simulations were done on an IBM 3033 AP at the Naval Postgraduate School using the LLRANDOMII random number generating package; Lewis and Uribe[17]. The simulations reported in Tables (5.2) and (5.3) have 500 replications. Generated in each replication is a sample of $n_u = 25$ up times and $n_d = 25$ down times. In the experiment reported in Table (5.2) the up times and down times are independent. Three distributions are used to generate the down times:

Table 5.2

Estimates of Availability
Independent Up and Down Times

Time	Dist	True A(t)	β_p Mean (Median)	β_p Var	$\hat{A}_p(t)$ Mean	$\hat{A}_p(t)$ MSE	β_e Mean (Median)	β_e Var	$\hat{A}_e(t)$ Mean	$\hat{A}_e(t)$ MSE	β_q Mean (Median)	β_q Var	$\hat{A}_q(t)$ Mean	$\hat{A}_q(t)$ MSE
.2	A	.83	3.1 (3.1)	.14	.85	.03	3.8 (3.4)	4.2	.83	.06	3.8 (3.4)	4.7	.83	.06
.5	(Gamma)	.71	3.1 (3.1)	.14	.74	.05	4.6 (3.6)	34	.72	.06	4.5 (3.7)	23	.72	.06
1.0		.67	3.1 (3.1)	.14	.68	.05	4.7 (3.6)	48	.68	.05	4.6 (3.5)	23	.68	.05
.2	B	.85	3.1 (3.1)	.23	.85	.03	3.4 (3.0)	3.2	.85	.05	3.3 (3.0)	3.1	.85	.06
.5	(Exp)	.74	3.1 (3.1)	.23	.74	.04	3.5 (3.0)	4.7	.74	.06	3.5 (3.0)	4.7	.74	.06
1.0		.68	3.1 (3.1)	.23	.68	.05	3.7 (3.0)	14	.69	.06	3.9 (3.0)	18	.69	.06
.2	C	.90	4.4 (3.4)	6.7	.86	.06	3.4 (1.7)	38	.91	.03	3.5 (1.8)	52	.91	.03
.5	(M. Exp.)	.87	4.4 (3.4)	6.7	.77	.13	2.9 (1.2)	57	.87	.04	2.5 (1.2)	14	.87	.04
1.0		.84	4.4 (3.4)	6.7	.72	.17	2.5 (1.0)	26	.82	.06	1.9 (.92)	6.1	.83	.06
2.0		.79	4.4 (3.4)	6.7	.70	.17	2.5 (.87)	75	.77	.09	1.8 (.80)	9.8	.78	.08

Table 5.3

Estimates of Availability
Dependent Up and Down Times

Time	Case	True A(t)	$\phi_U\phi_D$ β_q Mean	$\phi_U\phi_D$ β_q Var	$\phi_U\phi_D$ $\hat{A}_q(t)$ Mean	$\phi_U\phi_D$ $\hat{A}_q(t)$ MSE	ϕ_C β_q Mean	ϕ_C β_q Var	ϕ_C $\hat{A}_q(t)$ Mean	ϕ_C $\hat{A}_q(t)$ MSE
0.2	DA	.94	3.1	1.7	.85	.10	1.0	.13	.94	.02
0.5		.87	3.1	1.1	.75	.13	1.0	.07	.87	.03
1.0		.79	3.1	1.0	.69	.11	1.0	.07	.79	.03
2.0		.72	3.2	1.9	.67	.05	1.1	.09	.71	.02
0.2	DB	.97	3.1	14	.91	.07	.73	5.1	.97	.02
0.5		.93	2.4	11	.87	.07	.83	3.8	.93	.03
1.0		.88	2.0	7.2	.83	.07	.79	.56	.88	.05
2.0		.81	1.8	8.0	.78	.08	.83	.45	.81	.07

A: $P\{D > t\} = 1 - e^{-4t} - 4te^{-4t}$, $t > 0$; (Gamma)

B: $P\{D > t\} = e^{-2t}$, $t > 0$; (Exponential)

C: $P\{D > t\} = .9e^{-9t} + .1e^{-.25t}$, $t > 0$; (Mixed Exponential).

The up times are generated from an exponential distribution with unit mean. The theoretical values of $A(t)$ in each of these cases are

A: $A(t) = .67 + e^{-4.5t}[.33\cos(1.9t) + .26\sin(1.9t)]$; (5.3)

B: $A(t) = .67 + .33e^{-3t}$; (5.4)

C: $A(t) = .67 + .24e^{-.34t} + .09e^{-9.9t}$. (5.5)

For each replication of the simulations whose results are reported in Table (5.2) the three estimates of $A(t)$ are computed. The mean, and mean square error (MSE), of $\hat{A}_p(t)$, $\hat{A}_e(t)$ and $\hat{A}_q(t)$ are computed: i.e. the mean is

$$\bar{A}_e(t) = \frac{1}{500}\sum_{k=1}^{500} \hat{A}_e(k;t) \tag{5.6}$$

and

$$MSE = \frac{1}{500}\sum_{k=1}^{500} (\hat{A}_e(k;t) - A(t))^2 \tag{5.7}$$

where $\hat{A}_e(k;t)$ is the point estimate at t in the kth realization. Furthermore, the mean and variance of the estimates of $\hat{\beta}$ are displayed

5.3 Discussion of Tables.

Table (5.2) reports results for simulations with data sample size 25. Not surprisingly, the exponential estimator $\hat{A}_p(t)$ has means equal to the values of $A(t)$ and the smallest mean square error in the case B of exponential down times. The means of $\hat{A}_p(t)$ are somewhat different from the A(t)'s in case A (gamma down times) and quite different in case C (mixed exponential down times). The estimators $\hat{A}_e(t)$ and $\hat{A}_q(t)$ have means which are closer to the A(t)'s than those of $\hat{A}_p(t)$ in cases A and C. The means of $\hat{A}_e(t)$ and $\hat{A}_q(t)$ are a bit low for t = 1, 2 in case C. In cases A and B the values of the MSE of $\hat{A}_e(t)$ and $\hat{A}_q(t)$ indicate the greater variability of these estimators. In the exponential case B, the true value of β is 3; the estimator $\hat{\beta}_p$ has means closest to 3 for case B and much smaller variances compared to $\hat{\beta}_e$ and β_q. In the other two cases p(t) is not of the form $e^{-\beta t}$; the theoretical values of A(t) are given in (5.3) and (5.5).

In case A (gamma distributed down times) the single parameter, β, of exponential decay is 4.5. The means of the estimators, β_e and β_q are closer to this value than those of β_p. In case C of mixed exponential down times, the smallest parameter of exponential decay is 0.341. Once again the means of β_e and β_q are closer to this value than are those of β_p. The variances of β_q and β_e for the sample size of 25 can be large. In many cases the variance of β_q is less than that for β_e. Increasing the sample size decreases the variances, as is anticipated.

Table (5.3) summarizes a simulation study of the procedure of Section 4.2 in two cases in which the pairs (U_i, D_i) are independently and identically distributed with U_i and D_i perfectly dependent. The two cases are:

DA: $U_i = E_i$ and $D_i = \frac{1}{2}E_i$ with $\{E_i\}$ independent identically distributed exponentials with unit mean.

$$\text{DB:}\quad U_i = E_i \text{ and } D_i = \begin{cases} \frac{1}{9}E_i & \text{with probability } .9, \\ 4E_i & \text{with probability } .1 \end{cases}$$

with $\{E_i\}$ independent exponential random variables with unit mean. In case DA

$$A(t) = \frac{2}{3} + \frac{1}{3} e^{-t} . \tag{5.8}$$

In case DB

$$A(t) = \frac{2}{3} + 0.11e^{-t} + 0.22e^{-0.27t} . \tag{5.9}$$

In the simulations two estimators for $\hat{F}_C(s)$ are used. The first (incorrectly) assumes that $\{U_i\}$ and $\{D_i\}$ are independent and uses $\phi_U(s)\phi_D(s)$ to estimate $\hat{F}_C(s)$. The second does not assume independence, and instead estimates $\hat{F}_C(s)$ by

$$\phi_C(s) = \frac{1}{n_C} \sum_{i=1}^{n_C} e^{-s[u_i+d_i]} . \tag{5.10}$$

The table shows results for the estimators $\hat{A}_q(t)$ and $\hat{\beta}_q$. The sample size is 25 and there are 500 replications.

Table (5.3) informs us that the estimators may be noticeably sensitive to the choice of the estimator of $\hat{F}_C(s)$. If U_i and D_i are dependent, then using $\phi_U(s)\phi_D(s)$ to estimate $\hat{F}_C(s)$ can be quite misleading. As a result, it is suggested that if there is a possibility that U_i and D_i may be dependent, $\phi_C(s)$ is the more model-robust estimate of $\hat{F}_C(s)$, and hence of the desired availabilities.

Note that in case DA, A(t) has exactly the presumed form of (4.10). When ϕ_C is used in this case, the mean value of the $\hat{\beta}_q$'s equal the theoretical value of unity in all but the case t = 2.0. In case

DB, $p(t)$ is actually a mixture of two exponentials, one having rate unity and the other having rate 0.27; when ϕ_C is being used, the mean values of the $\hat{\beta}_q$'s fall between the two correct rates. Emphasis of either rate, and the corresponding probability, can be achieved by weighting.

5.4 Semi-Parametric Confidence Intervals Illustrated.

Table (5.4) shows bootstrap confidence intervals for $A_q(t)$ and β_q at various times. A single sample of 25 up times and 25 down times is generated by simulation. The up times are independent with unit-mean exponential distribution; they are independent of the down times. The down times are independent with the gamma distribution, case B. One hundred bootstrap replications were then carried out and $\hat{A}_q$ and $\hat{\beta}_q$ are computed for each replication. Both estimators of $\hat{F}_C(s)$, $\phi_U(s)\phi_D(s)$ and $\phi_C(s)$, are used. Table (5.4) exhibits the 5th, 50th, and 95th order statistics of the estimates which give 95% confidence intervals of the parameters.

All confidence intervals for β_q cover the true values for the particular sample utilized. The confidence intervals for β_q cover the exponential decay parameter of value 4.5, for the gamma case. The confidence intervals for $\hat{A}_q(t)$ are the same for both methods of estimating $\hat{F}_C(s)$ suggesting that not much is lost by using $\phi_C(s)$ to estimate $\hat{F}_C(s)$ in the independent case. As a result it is suggested that if there is a possibility that $\{U_i\}$ and $\{D_i\}$ may be dependent then $\phi_C(s)$ is the more robust estimate of $\hat{F}_C(s)$.

A similar bootstrap experiment was carried out with a single sample of 25 up times and 25 down times generated from model DA in which the up and down times are dependent. One hundred bootstrap

Table 5.4

Bootstrap Confidence Intervals
Gamma Distributed Down Times
Independent Up and Down Times

Time	True	$\phi_U\phi_D$						ϕ_C					
		$\hat{\beta}_q$			$\hat{A}_q(t)$			$\hat{\beta}_q$			$\hat{A}_q(t)$		
	A(t)	.05	.50	.95	.05	.50	.95	.05	.50	.95	.05	.50	.95
.2	.83	1.6	4.1	13.6	.71	.84	.92	1.6	3.9	11.8	.71	.84	.92
.5	.71	2.4	4.1	12.9	.61	.73	.82	2.4	3.9	9.8	.61	.73	.82
1	.67	2.3	3.4	16.3	.58	.71	.79	2.2	3.4	19.1	.58	.71	.79

replications were generated and $\hat{A}_q$ and $\hat{\beta}_q$ computed for each replication with both estimators of $\hat{F}_C(s)$. Once again the resulting confidence intervals for A(t) were very similar for both estimators of $\hat{F}_C(s)$.

Of course the above results are quite fragmentary, but seem useful and promising. Further sampling experiments and asymptotic analyses will shed more light on the behavior of the estimating procedures explored, and may well suggest alternations or replacements.

6. CONCLUDING COMMENTS.

This paper argues that probability models of the availability of various systems can be expressed in terms of Laplace transforms, and that the finite-time behavior of such systems can be inferred, nearly non-parametrically, from data. Our approach has been to invoke the empirical Laplace transforms and to utilize its easy direct interpretation (EROBS), in conjunction with a presumed approximate exponential rate of approach to steady state, to deduce availability at a fixed finite time after a known initial moment (FOBS).

The methods proposed are distinguished by their simplicity and moderate computer intensivity as well as by their lack of direct dependence upon probability models in "up" and "down" times selected from conventional families such as the Gamma. Simulations have been used to evaluate the procedures suggested, and to provide approximate confidence limits, either by asumptotics (utilizing the approximately Normal/Gaussian behavior of the empirical transform), or by a simple re-sampling, Efron's bootstrap[18]. In particular, we have examined the effect on inference quality of assuming the wrong joint probability model: one that falaciously assumes independence when dependence (between up and down times) is actually present.

No claim is made that the methods proposed here are the best available; in fact there are many alternatives. One is to analytically invert the empirical transform of availability possibly by use of the Stehfert algorithm[19], although competitors are available. Another is to bootstrap directly; the latter exercise involves re-sampling up and down times from the observed data, to reconstruct the sample path of the process, and to score 1 at time t if an up time covers t, otherwise score 0. Finally, A(t) is estimated by the proportion of (re)samples that count 1. Confidence limits are available from the basic bootstra technology. The direct bootstrap approach is being investigated by Lee[20]. Another option is to replace the distribution functions in the renewal equations for availability by their empirical counterparts and numerically solve the empirical renewal (Volterra-type) equations, wit subsequent bootstrap follow-ups to assess uncertainty. All such methods promise to be far more computationally intensive than our present approximate approaches. Their investigation has been deferred Application of our approximation procedure to infer the M/G/1 queue finite-time behavior is under way; Jacobs and Gaver[11].

References

1. Goyal, A., Lavenberg, S.S., and Trivedi, K.S., "Probabilistic modelling of computer system availability," Annals of Operations Research, 8, 1987; pp. 285-306.
2. Cox, D.R., "Some remarks on Markov processes in medical statistics," Int. Symp. on semi-Markov processes and their applications, 1986, ed. J. Janssen, Plenum, N.Y.
3. Jacobs, P.A., "Estimation of the probability of a long time to the first entrance to a state in a semi-Markov model," Naval Postgraduate School Technical Report NPS55-87-012, Sept. 1987.
4. Gaver, D.P., "Point process problems in reliability," Stochastic Point Processes: Statistical Analysis, Theory and Applications, 1972, ed. P.A.W. Lewis, John Wiley, N.Y.
5. Thomas, L.C., Jacobs, P.A., and Gaver, D.P., "Optimal inspection policies for standby systems," Communications in Statistics: Stochastic Models, 3, No. 2, 1987; pp. 259-274.
6. Gaver, D.P., and Chu, B.B., "Jackknife estimates of component and system availability," Technometrics, 21, No. 4, 1979; pp. 443-450.
7. Cox, D.R., "Some problems of statistical analysis connected with congestion," Proc. of the Symposium on Congestion Theory, 1965, ed. W.L. Smith and W.E. Wilkinson, U. of North Carolina Press, Chapel Hill, pp. 289-310.
8. Gross, D., and Harris, C.M., Fundamentals of Queueing Theory, 2nd Ed., John Wiley, 1985.

9. Ascher, H. and Feingold, H., Repairable Systems Reliability: Modeling, Inference, Misconceptions and Their Causes, Marcel Dekker, Inc., 1984.

10. Gaver, D.P., "Observing stochastic processes and approximate transform inversion," Operations Research, 15, 1966; pp. 444-459.

11. Gaver, D.P. and Jacobs, P.A., "On inference and transient respons for M/G/1 models," Teletraffic Analysis and Computer Performance Evaluation (ed. O.J. Boxma, J.W. Cohen, and H.C. Tijms) Elsevier Science Publishers B.V. (North-Holland), 1986.

12. Feller, W., An Introduction to Probability Theory and Its Applications, Vol. II, John Wiley, New York, 1966.

13. Odoni, A.R., and Roth, E., "An empirical investigation of the transient behavior of stationary queueing systems," Operations Research, 31, 1983; pp. 432-455.

14. Morse, P.M., "Stochastic processes of waiting lines," Operations Research, 3, 1955; pp. 255-261.

15. Morse, P.M., Queues, Inventories, and Maintenance, John Wiley and Sons, New York, 1958.

16. Keilson, J., Markov Chain Models--Rarity and Exponentiality, Springer-Verlag, New York, 1979.

17. Lewis, P.A.W., and Uribe, L., "The new Naval Postgraduate School random number package--LLRANDOM II," Naval Postgraduate School Technical Report NPS55-81-005, Monterey, 1981.

18. Efron, B. and Tibshirani, R., "Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy," Statistical Science, 1. 1986; pp. 54-77.

19. Stehfert, H., Algorithm 368. "Numerical inversion of Laplace transforms," Comm. Assoc. Comput. Mach., 13, 1970; pp. 47-49.

20. Lee, C.H., Thesis for degree of Master of Science in Operations Research, Naval Postgraduate School (in process), 1988.

DISTRIBUTION LIST

	NO. OF COPIES
Library (Code 0142) Naval Postgraduate School Monterey, CA 93943-5000	2
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Office of Research Administration (Code 012) Naval Postgraduate School Monterey, CA 93943-5000	1
Center for Naval Analyses 4401 Ford Avenue Alexandria, VA 22302-0268	1
Library (Code 55) Naval Postgraduate School Monterey, CA 93943-5000	1
Operations Research Center, Rm. E40-164 Massachusetts Institute of Technology Attn: R. C. Larson and J. F. Shapiro Cambridge, MA 02139	1
Koh Peng Kong OA Branch, DSO Ministry of Defense Blk 29 Middlesex Road SINGAPORE 1024	1
Arthur P. Hurter, Jr. Professor and Chairman Dept of Industrial Engineering & Mgmt Sciences Northwestern University Evanston, IL 60201-9990	1
Institute for Defense Analysis 1800 North Beauregard Alexandria, VA 22311	1